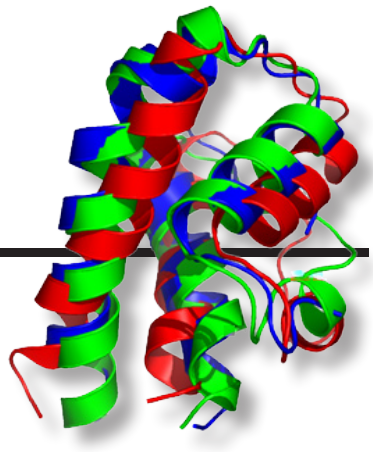
Argonne NATIONAL LABORATORY



High-Resolution Protein Structure Prediction

Proteins are the workhorse molecules of all biological systems. A deep and predictive understanding of life thus requires a detailed picture of their structure. Researchers from the University of Washington are testing a recent dramatic development in their ROSETTA methodology for determining protein structure computationally. Their advance allows for the prediction of full atomic-level structure of many proteins with unprecedented accuracy. This work will have an impact in the prediction of docking interfaces between proteins and the field of protein design.

A research group from the University of Washington is pushing the envelope of computational protein structure prediction using their ROSETTA methodology, which they have been developing over the past several years. Recent advances have led to the development of an all-atom protocol that can provide the full atomic-level structure of proteins. This new protocol has demonstrated successful prediction of protein structures to atomic-level accuracy in a community-wide test. This protocol uses a physically realistic high-resolution energy function that has been previously applied towards designing a novel protein fold with atomic-level accuracy. The ultimate goal of the project is to make available high-resolution structures of all proteins of biological interest. The results so far are an important milestone in this direction.

The ROSETTA method uses a two-phase Monte Carlo algorithm to sample the extremely large space of possible structures in order to find the

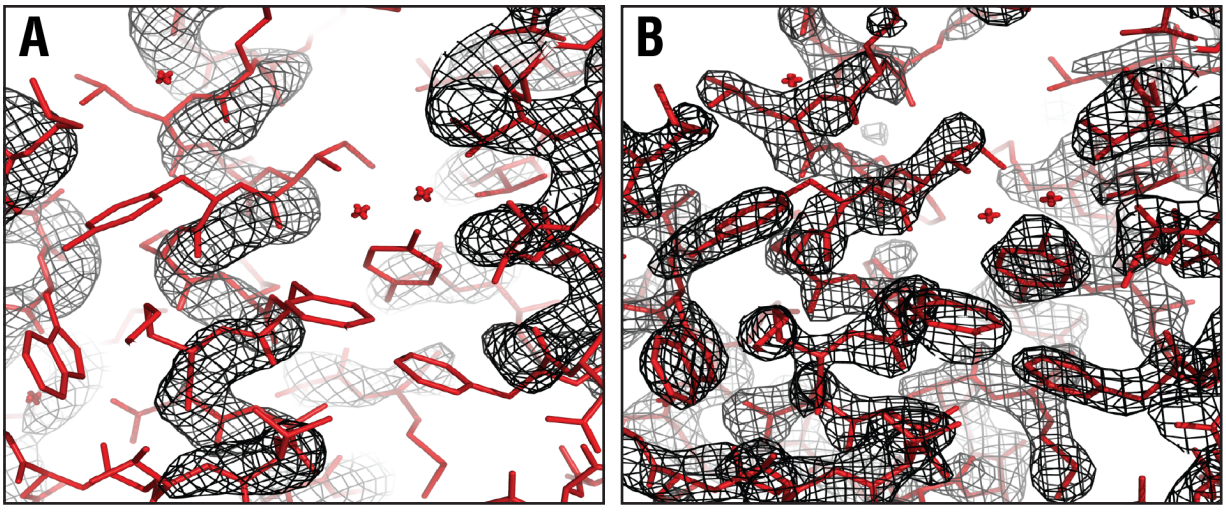


Figure 1. A picture of the electron density (black) generated from phases determined from the starting low-resolution NMR model (left panel) and that generated from phases determined from the Rosetta refined model (right panel). The crystal structure (red) agrees very well with the phases determined from the Rosetta model.

most favorable one. The first phase generates a low-resolution model of the protein backbone atoms while approximating the side chains with a single dummy atom. The high-resolution phase then uses a more realistic model of the full protein, along with the corresponding interactions, to find the best candidate for the native structure. Over the past year, the method has been tested on several proteins of known structure up to 189 amino acids in length. In many cases, the accuracy of the prediction was within a remarkable one angstrom of the experimentally solved high-resolution crystal structure.

Since the protein conformation search space is so large, such a high level of accuracy would not be possible without the availability of large amounts of computing power. Allocations of processor-hours from the Department of Energy's (DOE) INCITE program enabled the team to perform the extensive searches required by using the high-performance Blue Gene/L computers at IBM's T.J. Watson Research Center and the Argonne Leadership Computing Facility.

Currently, the method is being tested on proteins of known structure to determine the size range and fold classes for which accurate results can be predicted. Subsequent computations will focus on making predictions for proteins with unknown structures. Additional work will go into refining the method to expand the range of proteins to which it can be applied.

A key application of the ROSETTA method is in refining homology models or low-resolution models obtained from limited Nuclear Magnetic Resonance (NMR) data (Figs. 1, 2). A stringent test for the accuracy of the high-resolution protocol is its suitability for molecular replacement solutions used in phasing X-ray crystallographic data. Rosetta

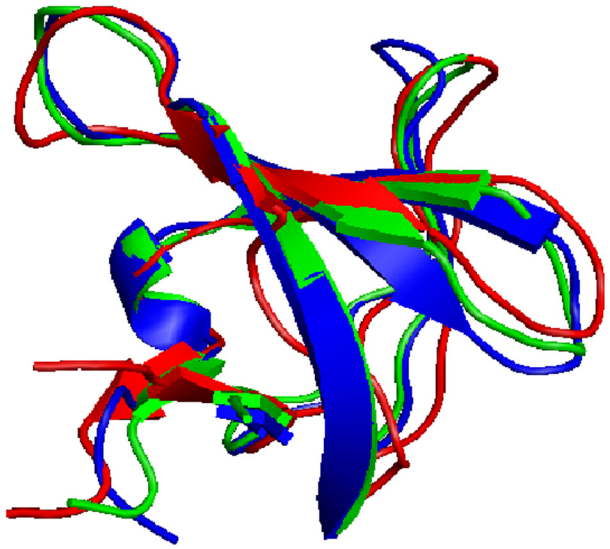


Figure 2. Native protein structure (blue) along with the starting low-resolution NMR model (red) and the Rosetta refined prediction (green).

refined models provide significantly improved molecular replacement solutions, paving the way for solving crystal structures in a high-throughput fashion. The availability of high-resolution protein structures would not only contribute towards basic science by helping researchers understand the molecular basis of protein function, it would greatly accelerate efforts towards structure-based drug design and protein engineering.

The Argonne Leadership Computing Facility and the INCITE program directly support the primary mission of DOE's Office of Advanced Scientific Computing Research to discover, develop, and deploy computational and networking tools that enable researchers in the scientific disciplines to analyze, model, simulate, and predict complex phenomena.

Contributors:

David Baker
University of Washington

Srivatsan Raman
University of Washington

For more information on this subject, please contact:

David Baker
Department of Biochemistry
University of Washington
206-543-1295
dabaker@u.washington.edu

For more information about ALCF, please contact:

Pete Beckman, Acting Division Director
Argonne Leadership Computing Facility
Argonne National Laboratory
630-252-9020
beckman@mcs.anl.gov

April 2008
alcf-bake-0408

UChicago ▸ Argonne LLC

A U.S. Department of Energy laboratory
managed by UChicago Argonne, LLC